Hindi Vidya Prachar Samiti's

Ramniranjan Jhunjhunwala College

of Arts, Science & Commerce

(Empowered Autonomous College)

Affiliated to

UNIVERSITY OF MUMBAI

Syllabus for the M.Sc.

Program: M.Sc. STATISTICS

Program Code: RJSPSTA

National Education Policy (NEP 2020)

Level 6.5

***(*CBCS 2025-2026)**

Credit Structure for M Sc. Semester III as per NEP 2020 to be Implemented from the academic year 2025-2026

Semester III

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPSTA301	Time Series Analysis	Major	3	-	3
RJSPSTA302	Machine Learning	Major	3	-	3
RJSPSTA303	Testing of Hypothesis	Major	3	-	3
RJSPSTA304	Database Management and Data Visualization	Major	2	-	2
RJSPSTAP301	Statistical Computing III (Practical based on Paper I, II and III in relevant software)	SC-1		6	3
RJSPSTAE301	Professional Elective-I(Data Analysis Using Statistical Softwares)	PE-1	4		4
RJSPSTAE302	Professional Elective-II (Big Data Technology)	PE-2	4	-	4
RJSPSTADM301	Data Mining	DM	4		4
	Total		**23**	**6**	**22**

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	III MAJOR (CORE) SUBJECT
TITLE OF THE SUBJECT/COURSE	:	TIME SERIES ANALYSIS
COURSE CODE	:	RJSPST301
CREDITS	:	03
DURATION	:	45 HOURS

LEARNING OBJECTIVES	
1	Students will be able to understand and analyze components of time series.
2	Students will be able to explain time series with different structures.
3	Students will be able to perform trend analysis, develop seasonal indices, analysis cyclical variations and random components
4	Students will be able to become an expert in time series analysis and applications.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Get cutting edge of knowledge in time series analysis and applications.	1,2	Understand
CO2	Demonstrate skills and knowledge in communicating statistical techniques for scientific data analysis.	1, 3,5	Apply and Analyze
CO3	Develop the ability to perform diagnostic checks and validate time series models for reliability.	1,3,5	Understand and Apply
CO4	Acquire the basic required knowledge to do research and to take up advanced studies in time series analysis	1,3	Understand and Apply

M. Sc Statistics Syllabus Semester III & IV

SEMESTER-III

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPSTA301	Time Series Analysis	DSC	Lectures	Practical	
			3	-	3

Unit	Topic	Lectures
I	Time-series as a discrete parameter stochastic process. Exploratory Time Series Analysis: Tests for trend and seasonality, Stationarity, Tests for Stationarity, Exponential and Moving average smoothing. Holt Winters smoothing. Forecasting based on smoothing, adaptive smoothing.	15
II	Probability Models: White noise model, random walk, linear processes, Moving Average (MA), Autoregressive (AR), ARMA and ARIMA models (with and without seasonal effect). SARIMA, Invertibility. ACF and PACF of these processes.	15
III	Significance of Autocorrelation, Sample ACF and PACF for Model Identification. Box-Jenkins forecasting, Estimation of ARIMA models parameters. Forecasting, Yule-Walker equations.Residual analysis and diagnostic checking. ARIMAX, ARCH and GARCH models, Properties, Examples, Estimation and forecasting.	15

References:

1. Anderson, T. W (1971): The Statistical Analysis of Time Series, Wiley, N.Y.

2. Brockwell, P.J. and Davis, R. A. Time-Series(1991): Theory and Methods (Second Edition), Springer-Verlag. 19

3. Box, G.E.P. and Jenkins, G.M. (1976): Time Series Analysis-Forecasting and control Hodlen-day, San Franciscor.

4. Kendall, Sir Maurice and Ord. J. K. (1990): Time Series (Third Edition) Edward Arnold.

5. Montgomery, D. C. and Johnson, L. A. (1977): Forecasting and Time Series Analysis, McGraw Hill.

6. Robert.H.Shumway and David.S.Stoffer (2017) : Time Series and its Applications (Fourth Edition) Springer

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	III MAJOR (CORE) SUBJECT
TITLE OF THE SUBJECT/COURSE	:	MACHINE LEARNING
COURSE CODE	:	RJSPST302
CREDITS	:	03
DURATION	:	45 HOURS

LEARNING OBJECTIVES	
1	Upon completion of this course, the student should be able to understand the basic theory underlying machine learning.
2	Students will be able to formulate machine learning problems corresponding to different applications.
3	Students will be able to understand a range of machine learning algorithms along with their strengths and weaknesses.
4	Students will be able to apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Apply structured thinking to unstructured problems.	1,3,5	Apply
CO2	Understand a very broad collection of machine learning algorithms and problems .	1, 3	Understand
CO3	Learn algorithmic topics of machine learning and mathematically deep enough to introduce the required theory.	1,3	Remember and Understand
CO4	Develop an appreciation for what is involved in learning from data.	2,3,6	Apply

Semester-III

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPSTA302	Machine Learning	Major	Lectures	Practical	
			3	-	3

Unit	Topic	Lectures
I	**Introduction to machine learning:** How machine learn, Basic structure of Machine Learning Model and comparison to traditional programming, Goal and Applications of Machine learning. Types of Data related to Machine Learning,Introduction to standard machine learning datasets and interpreting the data,Correlation between data,supervised vs Unsupervised learning algorithms. **Feature Engineering:**Feature transformation, feature construction, feature scaling(Standardization & Normalization), feature extraction: Principal Component Analysis, t-sne, curse of dimensionality, function and power transformation.	15
II	**Regression :** Linear regression and assumptions, Feature Selection using Correlation Coefficient, Polynomial regression, Multivariate Regression, Logistic regression, Cost functions for Regression(Mean Square Error, Mean Absolute Error, Root Mean Square error, R2 Score) and optimization using gradient descent. **Classification :** Linear & Non-Linear Datasets, Decision Tree Algorithm, Role of Entropy and Information Gain(IG) in decision tree algorithm, CART ,Random forest, Classification Metrics - Type I and Type II Errors and Confusion Matrix, Overfitting vs Underfitting, Bias variance tradeoff, Naive Bayes Classifier, Support Vector Machine (SVM), Kernels in SVM.	15
III	**Introduction to Clustering :** Evaluation the output of clustering methods, Partition Based Clustering,K-Means clustering, elbow methods to find number of clusters, K-Medoids clustering, Hierarchical Clustering, Agglomerative and Divisive Approach, Dendrogram, linkage, Density-based spatial clustering of applications with noise(DBSCAN).	15

References:

1. "An Introduction to Statistical Learning With Application in R", By Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer Texts in Statistics.

2. "Machine Learning", Mitchell Tom, McGraw Hill,1997.

3. "Pattern classification", 2nd edition, Richard O. Duda, Peter E. Hart, David G. Stork. Wiley, New York, 2001.

4. "Machine Learning: A Probabilistic Perspective", Kevin P. Murphy, MIT Press, 2012.

5. "Practical Data Science", Andreas Francois Vermeulen, APress, 2018.

6. "Principles of Data Science", Sinan Ozdemir, Packt, 2016.

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	III MAJOR (CORE) SUBJECT
TITLE OF THE SUBJECT/COURSE	:	TESTING OF HYPOTHESIS
COURSE CODE	:	RJSPSTA303
CREDITS	:	03
DURATION	:	45 HOURS

LEARNING OBJECTIVES	
1	Upon completion of this course, the student should be able to understand key concepts and terminology related to hypothesis testing, such as null and alternative hypotheses, Type I and II errors, significance levels, and p-values.
2	Students will be able to understand and explain the fundamental principles and procedures of hypothesis testing.
3	Upon completion of this course, the student should be able to understand how different parametric and non parametric tests are used to solve real life problems.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Apply suitable hypothesis testing methods to analyze and interpret data accurately.	1,3,5	Apply draw conclusions
CO2	Construct different types of tests for testing different hypotheses.	1, 3,5	Understand and Apply
CO3	Evaluate the appropriateness of different hypothesis tests based on data and assumptions.	1,3,5,6	Apply, analyze and evaluate
CO4	Understand when to use non parametric tests and apply them appropriately	1,3	Understand and Apply

SEMESTER-III

M. Sc Statistics Syllabus Semester III & IV

Course code	**Course name**	**Group**	**Teaching Scheme (Hrs/Week)**		**credits**
RJSPSTA303	Testing of Hypothesis	Major	Lectures	Practical	
			3	-	3

Unit	Topic	Lectures
I	Introduction of hypothesis testing, test function, MP tests, Neyman- Pearson lemma, UMP tests, nonexistence of UMP tests.	15
II	Unbiased test, LMP test, α- Similar test, Likelihood Ratio Test.	15
III	Goodness of fit tests: Chi-square goodness of fit test, The Kolmogorov- Smirnov one statistic test. Sign test, Wilcoxon Signed-rank test.Wald Wolfowitz run test, Kolmogorov-Smirnov two sample test, Mann-Whitney U-test, Wilcoxon Rank-Sum test.Measures of association and their tests of significance: Kendall's Tau coefficient, Spearman's coefficient of rank correlation.The Kruskal-Wallis Test,Friedman's test.	15

References:

1. Vijay K. Rohatgi, A. K. M. D. Ehsanes Saleh: An introduction to probability and statistics. John Wiley and Sons:- (2nd Edition)
2. J. D. Gibbons & S. Chakrabarti : Nonparametric Statistical Inference. (3rd Edition, Revised and Expanded).
3. Lehmann, E.L. & Romano Joseph P. (2005) : Testing Statistical Hypotheses (3rd Edition):- Springer Text.
4. Casella, G. and Berger, R. L. (2002). Statistical Inference. Duxbury Advanced Series, Second Edition.
5.Lehmann, E. L. and Romano, J. (2005). Testing Statistical Hypotheses, Springer.
6. Shao, J. (2003). Mathematical Statistics, Springer-Verlag, New York.

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	III MAJOR (CORE) SUBJECT
TITLE OF THE SUBJECT/COURSE	:	DATABASE MANAGEMENT AND DATA VISUALISATION
COURSE CODE	:	RJSPSTA304
CREDITS	:	02
DURATION	:	30 HOURS

LEARNING OBJECTIVES	
1	Students will be able to understand database concepts such as tables, records, primary keys, foreign keys, and relationships.
2	Students will be able to explain the structure and functions of database management systems (DBMS).
3	Upon completing this course students will be able to use SQL queries to create, read, update, and delete data in a database.
4	Explain principles of effective data visualization and visual perception.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Define and explain core concepts and architecture of database management systems.	1,3	Remember
CO2	Apply SQL commands to manipulate and query databases effectively.	1, 3	Understand and Apply
CO3	Analyze data to select suitable visualization techniques for different data types.	1,3	Analyse
CO4	Design and develop innovative visualizations to represent complex data stories.	1,3,5	Apply

Course code	**Course name**	**Group**	**Teaching Scheme**	**Credits**

			(Hrs/Week)		
RJSPSTA304	Database Management and Data Visualisation	Major	Lectures	Practical	
			2	-	2

Unit	Topic	Lectures
1	Components of ER Diagram, Converting ER Diagram to Relational Schema,Data Definition Language (DDL):Creating and Managing Tables Constraints ,Views , Data Manipulation Language (DML),Inserting Data ,Updating Data ,Deleting Data , Data Retrieval Language (DRL),Retrieving Data using Filtering ,Sorting,Joins (INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL JOIN, Data Control Language (DCL),Managing User Permissions ,Transaction Control.	15
2	Overview of Business Intelligence & Data Visualization, Introduction to Tableau: Features & Architecture,Installing & Setting up Tableau Desktop, Connecting to Data Sources,Importing data from Excel, SQL, and Cloud Sources,Data Extraction vs. Live Connection,Data Cleaning & Preprocessing in Tableau,Data Visualization in Tableau,Understanding Worksheets, Dashboards, and Stories,Creating Basic Charts: Bar, Line, Pie, Scatter, and Maps,Filters, Parameters, and Hierarchies Advanced Visualization & Calculations,Dual-axis & Combination Charts,Table Calculations & Quick Table Calculation,Using Calculated Fields & LOD (Level of Detail) Expressions, Dashboard & Report Creation,Designing Interactive Dashboard,Adding Actions (Filters, Highlights, URL Links),Formatting & Customizing Reports	15

References

1. The Big Book of Dashboards by Steve Wexler, Jeffrey Shaffer, Andy Cotgreave, John Wiley & Sons, 2017 4. Practical Tableau by Ryan Sleeper, O"Reilly Media, 2018
2. Power BI. Book-1, Business Intelligence Clinic: Create and Learn by Roger F Silva, 2018
3. Introducing Microsoft Power BI by Alberto Ferrari and Marco Russo, Microsoft Press, Washington, 2016

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	III PROFESSIONAL ELECTIVE -I
TITLE OF THE SUBJECT/COURSE	:	DATA ANALYSIS USING STATISTICAL SOFTWARE
COURSE CODE	:	RJSPSTAE301
CREDITS	:	04
DURATION	:	60 HOURS

LEARNING OBJECTIVES	
1	Students will be able to understand how statistical software is used for analyzing data.
2	Students will be able to understand how these software give quick results of any problem related to statistical modeling.
3	Upon completing this course students will be able to code and run problems easily.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Use python libraries for data visualization	1,3,4	Understand
CO2	Conduct exploratory data analysis using Python and Excel.	1, 3	Apply and analyze
CO3	Interpret results of exploratory data analysis.	1,3	Analyze and Evaluate
CO4	Paraphrase the results for documentation	1,2	

SEMESTER-III

M. Sc Statistics Syllabus Semester III & IV

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		credits
RJSPSTAE301	Professional Elective-I(Data Analysis Using Statistical Softwares)	Elective	Lectures	Practical	
			4	-	4

Unit	Topic	Lectures
I	Review of Excel, sorting, filtering and construction of charts. Entering Excel Formulas and Formatting Data. Creating Excel Functions, Filling Cells, and Printing. More on Entering Excel Formulas: SUMIF Function, AVERAGEIF Function, COUNTIF Function, IF Function, AND Function, FREQUENCY Function. Matrix operations- Transpose, Product and Inverse operations using Excel. Pivot tables and look up functions.	15
II	Data Analysis Pack in Excel, calculating theoretical probability using Binomial, Poisson and Normal distributions., descriptive statistics, tests of hypothesis, ANOVA, Correlation and Regression, Random Number Generation from different distributions, Binomial, Poisson, Uniform, Normal and from discrete distributions with given mean and variance. Forecasting Using Excel – Moving Averages and Exponential Smoothing, Use of functions, Linest, Logest, Forecast , Growth, Trend for trend analysis.	15
III	Measure of central tendency and Measures of dispersion using Python, Advanced Concepts in Python: Writing python program for: Testing of Hypothesis, testing One-sample and two-sample hypothesis testing. Statistical tests using Python: t - test, z - test, Chi - square test, ANOVA.	15
IV	Karl Pearson correlation coefficient and Rank Correlation coefficient using Python, Linear regression - Simple and Multiple, Selection of variables in Multiple Linear Regression – stepwise procedures and analysis of residuals. Procedure for Binary Logistic regression, Factor analysis, Linear Discriminant analysis and Cluster analysis.	15

References:

1. Excel Charts, by John Walkenbach
2. Quick Start Guide from Beginner to Expert, by William Fischer
3. Excel 2016 Bible, by John Walkenbach
4. Nelli, F. (2015). Python Data Analytics: Data Analysis and Science using pandas.
5. Idris, I. (2016). Python Data Analysis Cookbook.
6. Phuong, Vo. T. H. and Czygan, M. (2015). Getting Started with Python Data Analysis.

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	III PROFESSIONAL ELECTIVE -II
TITLE OF THE SUBJECT/COURSE	:	BIG DATA TECHNOLOGY
COURSE CODE	:	RJSPSTE302
CREDITS	:	04
DURATION	:	60 HOURS

LEARNING OBJECTIVES	
1	Students will be able to understand basic and advanced methods of big data technology and tools.
2	Students will be able to understand the knowledge of MapReduce, Hadoop and its ecosystem.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Understand build and maintain, reliable, scalable and distributed systems with Apache Hadoop.	1	Understand
CO2	Access and Process Data on Distributed File System	1, 3	Apply, analyze and evaluate
CO3	Develop Big Data Solutions using Hadoop Eco System and Analyze Infosphere BigInsights Big Data Recommendations.	1,3	Apply, analyze and evaluate
CO4	Identify Big Data and its Business Implications and List the components of Hadoop and Hadoop Eco-System.	1,2	Apply, analyze and evaluate

SEMESTER-III

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPSTE302	Professional Elective-II (Big Data Technology)	Elective	Lectures	Practical	
			4	-	4

Unit	Topic	Lectures
I	INTRODUCTION TO BIG DATA Introduction: Distributed file system, Big Data and its importance, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce. INTRODUCTION HADOOP Big Data: Apache Hadoop & Hadoop Ecosystem, Moving Data in and out of Hadoop, Understanding inputs and outputs of MapReduce, Data Serialization.	15
II	HADOOP ARCHITECTURE Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands, Anatomy of File Write and Read., Name Node, Secondary Name Node, and Data Node, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers, Cluster Setup, SSH & Hadoop Configuration, HDFS Administering, Monitoring & Maintenance.	15
III	HADOOP ECOSYSTEM AND YARN Hadoop ecosystem components: Schedulers, Fair and Capacity, Hadoop 2.0 New Features Name Node High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN.	15
IV	Hive and HiveQL, HBase Hive Architecture and Installation, Comparison with Traditional Database. HiveQL Querying Data, Sorting and Aggregating, Map Reduce Scripts, Joins & Subqueries. HBase concepts Advanced Usage, Schema Design, Advance Indexing, PIG, Zookeeper, how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with Zookeeper.	15

References

1. "Professional Hadoop Solutions", Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Willey, 2015.
2. "Understanding Big data", Chris Eaton, Dirk deroos et al, McGraw Hill, 2012
3. "HADOOP: The definitive Guide", Tom White, O Reilly 2012. 6 IT2015 SRM(E&T).
4. "Big Data Analytics with R and Hadoop", Vignesh Prajapati, Packet Publishing 2013.
5. "Oracle Big Data Handbook", Tom Plunkett, Brian Macdonald et al, Oracle Press, 2014.
6. "Big Data and Business analytics", Jy Liebowitz, CRC press, 2013.
7. http://www.bigdatauniversity.com/

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	III (Practical Based on Major Paper-I ,II and III)
TITLE OF THE SUBJECT/COURSE	:	Statistical Computing-III
COURSE CODE	:	RJSPSTAP301
CREDITS	:	03
DURATION	:	90 HOURS

LEARNING OBJECTIVES	
1	Students will be able to apply various time series models (e.g., AR, MA, ARIMA) to real datasets using software tools.
2	Students will be able to understand the process of training, testing, and evaluating ML models.
3	Students will be able to apply hypothesis tests on sample data using software tools..

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Build and fit appropriate time series models to given datasets.	1, 2,3,5	Understand and apply
CO2	Apply classification, regression, and clustering algorithms on datasets.	1, 2,3,5	Apply
CO3	Interpret results and draw valid conclusions based on test outcomes..	1, 2,3,5	Analyze and Evaluate

SEMESTER-III

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		credits
RJSPSTAPA301	Statistical Computing-III	Practical	Lectures	Practical	
			-	6	3

List of Practical's:

Practical 1: Time Series Analysis -I
Practical 2: Time Series Analysis-II
Practical 3: Time Series Analysis-III
Practical 4: Implementation of regression models
Practical 5: Perform all feature engineering operations on the dataset
Practical 6: Implementation of binary classifier.
Practical 7: Implementation of multi-class classification and Distance based models.
Practical 8: Perform regularization in machine learning.
Practical 9: Implementation of Hierarchical Clustering.
Practical 10: Implementation of K-Means clustering and elbow method for optimal number of clusters.
Practical 11: MP and UMP test
Practical 12: LRT Test
Practical 13: Non Parametric Test

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	III
TITLE OF THE SUBJECT/COURSE	:	Data Mining
COURSE CODE	:	RJSPSTADM301
CREDITS	:	04
DURATION	:	60 HOURS

LEARNING OBJECTIVES	
1	Students will be able to understand key concepts of data mining such as data preprocessing, classification, clustering, association rules, and pattern discovery.
2	Students will be able to use data mining algorithms to extract meaningful patterns from large datasets.
3	Students will be able to analyze data mining results to identify trends, patterns, and relationships.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Apply various data mining algorithms for classification, clustering, and association rule mining on real datasets.	1,2,5	Apply
CO2	Analyze and interpret the patterns and knowledge discovered through data mining processes.	1,2,5	Analyze
CO3	Develop data mining models and implement solutions for practical data analysis problems using relevant software/tools.	1,2,5	Apply

SEMESTER III

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPSTADM301	Data Mining		Lectures	Practical	
			4	-	4

Unit	Topic	Lectures
I	Data Mining—On What Kind of Data? Relational Databases, Data Warehouses, Transactional Databases, Advanced Data and Information Systems and Advanced Applications. Data Mining Functionalities—What Kinds of Patterns Can Be Mined? Concept/Class Description: Characterization and Discrimination, Mining Frequent Patterns, Associations, and Correlations, Classification and Prediction, Cluster Analysis, Outlier Analysis, Evolution Analysis.	15
II	Data Pre-processing : Descriptive Data Summarization , Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization and Concept Hierarchy Generation .	15
III	Data Warehouse and OLAP Technology: What is a Data Warehouse? Data Warehouse Architecture, Data Warehouse Implementation, Data warehouse and data mining.	15
IV	Classification and Prediction: What Is Classification?, What Is Prediction?, Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Rule-Based Classification, Support Vector Machines, Lazy Learners (or Learning from Your Neighbours).	15

References

1. Jiawei Han University of Illinois at Urbana-Champaign Micheline Kamber : Data Mining: Concepts and Techniques (Second Edition).

2. Ian H.Witten, Eibe Frank, Morgan Kaufman: Data Mining – Practical Machine Learning Tools and Techniques.

3. GalitShmueli, Nitin Patel, Peter Bruce, (2010): Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner , Wiley.

4. David Hand, HeikkiMannila and Padhraic Smyth: "Principles of Data Mining", MIT PRESS.

Credit Structure for M Sc. Semester IV as per NEP 2020 to be Implemented from the academic year 2025-2026

Semester IV

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPSTA401	Statistical Process and Quality Control	DSC	4	-	4
RJSPSTA402	Organizational Behaviour	DSC	2		2
RJSPSTAP401	Statistical Computing IV (Practical based on Paper I)	Practical	-	4	2
RJSPSTAE401	Professional Elective-I(Six Sigma and Lean: Foundation and Principles)	PE-1	4		4
RJSPSTAE402	Professional Elective-II(Data Warehousing)	PE-2	4	-	4
RJSPSTARP401	Research Project	RP		10	10
	Total				**22**

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	IV MAJOR (CORE) SUBJECT
TITLE OF THE SUBJECT/COURSE	:	STATISTICAL PROCESS AND QUALITY CONTROL
COURSE CODE	:	RJSPSTA401
CREDITS	:	04
DURATION	:	HOURS

LEARNING OBJECTIVES	
1	Upon completion of this course, the student should be able to understand key concepts such as control charts, process capability, assignable and common causes of variation.
2	Students will be able to fConstruct and interpret various control charts for variables and attributes.
3	Students will be able to differentiate between capable and non-capable processes using process capability indices (Cp, Cpk).
4	Students will be able to design quality control procedures and control chart systems for given processes.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Recall essential terms and techniques used in SPC and quality control.	1,2,3	Remember
CO2	Explain the working principles of control charts and sampling plans in maintaining quality.	1,3,5	Understand
CO3	Evaluate the effectiveness of SPC tools in detecting variations and maintaining process stability.	1,3,6	Remember and Understand
CO4	Design and implement a quality control system tailored to specific industrial or organizational needs.	2,3,6	Apply and Evaluate

Semester IV

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPSTA401	Statistical Process and Quality Control	Major	Lectures	Practical	
			4	-	4

Unit	Topic	Lectures
I	General Theory of Control Charts: Process and Product Control, control charts, 3-σ Control Limits, Tools for S.Q.C., Control Charts for Variables: $\overline{X}$ and R Charts, Criterion for Detecting Lack of Control in $\overline{X}$ and R Charts, Interpretation of $\overline{X}$ and R Charts, Control Chart for Standard Deviation or σ-chart. Control chart for Attributes: p-chart or Control chart for fraction defective, Control chart for Number of defective (d-chart), Interpretation of p-Chart, Control chart for number of Defects per unit(c-chart), Limits for c-chart, Application of c-chart.	15
II	Natural Tolerance Limits and Specification Limits, Modified Control Limits, Acceptance sampling by Attributes, Acceptance Quality Level (A.Q.L.) , Lot Tolerance Percentage Defective, Process Average Fraction Defective Defective.	15
III	Consumer's Risk, Producers Risk, Rectifying Inspection Plans, Average Outgoing Quality(A.O.Q), Average Outgoing Quality Limit(A.O.Q.L), Average Sample Number.	15
IV	Dodge and Roming Sampling Plans, Single Sampling Plan, Determination of n and c, A.O.Q.L. for double sampling plan, O.C. Curve, Double Sampling plan , O.C. of double Sampling Plan Consumer and Producers risk for Double Sampling Plan, A.S.N. Curve for Double Sampling Plan, Single V/S Double Sampling Plan.	15

References

1. Montgomery, D.C., Jennings, C.L. & Kulahci, M. (2015). Introduction to Time Series Analysis and Forecasting, 2nd Edition. Wiley.
2. Brockwell, P.J. & Davis R.A. (2016). Introduction to Time Series and Forecasting, 2nd Edition. Springer.
3. Montgomery, D.C. (2012). Introduction to Statistical Quality Control, 7th Edition. Wiley.
4. Grant, E. & Leavenworth, R. (2012). Statistical Quality Control, 7th Edition. Tata McGraw Hill.

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	IV
TITLE OF THE SUBJECT/COURSE	:	ORGANIZATIONAL BEHAVIOUR
COURSE CODE	:	RJSPST402
CREDITS	:	02
DURATION	:	30 HOURS

LEARNING OBJECTIVES	
1	Students will be able to understand key concepts such as personality, motivation, leadership, group dynamics, organizational culture, and conflict.
2	Students will be able to explain how individual behavior impacts group and organizational performance.
3	Students will be able to apply motivational theories and leadership styles to real-world workplace situations.
4	Students will be able to analyze workplace scenarios to identify behavioral issues at individual, group, or organizational levels.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Identify and define core concepts and theories of organizational behaviour.	1,2	Remember
CO2	Describe how individuals and groups behave in organizational settings and why.	1, 3,5	Understand
CO3	Apply OB principles to solve common organizational issues related to motivation, leadership, and communication.	1,3,5	Understand and Apply
CO4	Critically evaluate organizational policies and practices from a behavioral perspective.	1,3	Analyze

SEMESTER IV

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPSTOB401	Organizational Behaviour	Major	Lectures	Practical	
			2	-	2

Unit	Topic	Lectures
I	Organizational Behaviour I • Introduction to Organizational Behaviour-Concept, definitions, Evolution of OB • Importance of Organizational Behaviour-Cross Cultural Dynamics, Creating Ethical Organizational Culture & Climate • Individual and Group Behaviour-OB models–Autocratic, Custodial, Supportive, Collegial & SOBC in context with Indian OB • Human Relations and Organizational Behaviour Organisational Behaviour II • Managing Communication: Conflict management techniques. • Time management strategies. • Learning Organization and Organizational Design • Rewards and Punishments-Termination, layoffs, Attrition, Retrenchment, Separations, Downsizing	15
II	Resume Making and Building Interview Preparation and Mock Interview	15

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	IV (Professional Elective-I)
TITLE OF THE SUBJECT/COURSE	:	SIX SIGMA AND LEAN: FOUNDATION AND PRINCIPAL
COURSE CODE	:	RJSPSTAE401
CREDITS	:	04
DURATION	:	60 HOURS

LEARNING OBJECTIVES	
1	Students will be able to understand the fundamental principles and philosophy of Six Sigma.
2	Students will learn the DMAIC (Define, Measure, Analyze, Improve, Control) methodology for process improvement
3	Students will be able to Identify and analyze process inefficiencies and root causes of defects.
4	Students will be able to develop skills to collect, interpret, and use data to support decision-making.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Understand and apply the DMAIC (Define, Measure, Analyze, Improve, Control) methodology to process improvement scenarios.	1,2,3	Understand and Apply
CO2	Use basic Six Sigma tools such as SIPOC diagrams, Pareto charts, cause-and-effect diagrams, and control charts for quality analysis.	1, 3,5	Apply
CO3	Demonstrate awareness of Six Sigma roles (e.g., Green Belt, Black Belt) and how Six Sigma integrates with other quality systems like Lean or TQM.	1,3,5	Understand and Apply

SEMESTER-IV

Professional Elective-I

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		credits
RJSPSTAE401	Six Sigma and Lean: Foundation and Principles	Elective	Lectures	Practical	
			4	-	4

Unit	Topic	Lectures
I	Introduction to Lean and six - sigma: Quality definition, Definition of Lean, 5 S in Lean, 7 wastes in lean, 5 principles of lean. Definition of six - sigma and definition of Lean six - sigma. Measuring current performance (Sigma Level)	15
II	DMAIC overview, Define phase: VOC, VOB, VOP, CTQ, COPQ, Process mapping, Problem Statement Goal Statement, Project charter, DPU, DPMO, Yield, Brain Storming,Negative Brain storming, SIPOC, Cause and Effect diagram, Pareto chart, Run chart, Tier chart, Box-Whisker chart, Multivariate chart and FMEA	15
III	Measure phase: Process definition, Value Stream Mapping, MSA, Process Capability Analysis, statistical techniques: Data types, Data collection, Data measurement scale, Averages, Dispersion. Analyze Phase: Root Cause analysis, Correlation and Regression (Simple and Multiple), Probability distributions, Determination of sample size, Testing of Hypothesis(Parametric and Non- parametric)	15
IV	Improve Phase: Multi voting, Delphi Technique, Nominal group technique, Kaizen, Prioritization Matrix, Pug Matrix, SCAMPER. Control Phase: Control plans, Poke Yoke, SPC: IMR chart, X-bar, R-charts, P-chart, C and U charts	15

References

George, M. L, Maxey, J., Rowlands, D. T. & Upton, M. (2004). Lean six sigma pocket toolbook. McGraw-Hill Professional Publishing.
1. Gygi, C., & Williams, B. (2012). Six sigma for dummies. John Wiley & Sons.
2. Brussee, W. (2012). Statistics for Six Sigma Made Easy! Revised and Expanded Second Edition. McGraw Hill Professional.

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	IV (Professional Elective-II)
TITLE OF THE SUBJECT/COURSE	:	DATA WAREHOUSING
COURSE CODE	:	RJSPSTAE402
CREDITS	:	04
DURATION	:	60 HOURS

LEARNING OBJECTIVES	
1	Students will be able to understand Perform ETL process on source data and send it to the data warehouse database.
2	Students will learn how data warehousing supports business intelligence and decision-making.
3	Students will learn how to design and implement a basic data warehouse.
4	Students will explore OLAP (Online Analytical Processing) operations and their role in data analysis.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Describe the architecture, components, and lifecycle of a data warehouse system.	1,2,3	Understand and Apply
CO2	Design dimensional models using star and snowflake schemas for analytical processing.	1, 3,5	Apply
CO3	Use data warehousing concepts in modern technologies such as cloud-based data platforms and big data analytics tools.	1,5	Understand
CO4	Implement ETL processes to extract, clean, and load data into a data warehouse.	1,3,5	Understand and Apply

SEMESTER-IV

Professional Elective-II

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPSTAE402	Professional Elective-II(Data Warehousing)	**Elective**	**Lectures**	**Practical**	
			4	**-**	**4**

Unit	Topic	Lectures
I	Data Warehousing, Business Intelligence, and Dimensional Modeling Primer Different Worlds of Data Capture and Data Analysis, Goals of Data Warehousing and Business Intelligence, Dimensional Modeling Introduction, Kimball's DW/BI Architecture, Alternative DW/BI Architectures, Dimensional Modeling Myths.	15
II	Kimball Dimensional Modeling Techniques Overview Fundamental Concepts, Basic Fact Table Techniques, Basic Dimension Table Techniques, Integration via Conformed Dimensions, Dealing with Slowly Changing Dimension Attributes, Dealing with Dimension Hierarchies.	15
III	Retail Sales Four-Step Dimensional Design Process, Retail Case Study, Dimension Table Details, Retail Schema in Action, Retail Schema Extensibility, Fact less Fact Tables, Dimension and Fact Table Keys, Resisting Normalization Urges. Order Management Order Management Bus Matrix, Order Transactions, Invoice Transactions, Accumulating Snapshot for Order Fulfilment Pipeline.	15
IV	Inventory Value Chain Introduction, Inventory Models, Fact Table Types, Value Chain Integration, Enterprise Data Warehouse Bus Architecture, Conformed Dimensions. Customer Relationship Management Overview, Customer Dimension Attributes, Bridge Tables for Multivalued Dimensions, Complex Customer Behaviour, Customer Data Integration Approaches, Low Latency Reality Check.	15

References

1. "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling", Ralph Kimball Margy Ross, Wiley.
2. "The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data", Ralph Kimball, Joe Caserta.
3. "Building the Data Warehouse", Fourth Edition, W. H. Inmon, Wiley.

COURSE OUTCOMES (COs) M. Sc. Statistics

SEMESTER	:	IV (Practical Based on Major Paper-I)
TITLE OF THE SUBJECT/COURSE	:	Statistical Computing-IV
COURSE CODE	:	RJSPSTAP401
CREDITS	:	02
DURATION	:	60 HOURS

LEARNING OBJECTIVES	
1	Students will be able to learn to construct and interpret various control charts (e.g., $\bar{X}$-R, $\bar{X}$-S, p, np, c, and u charts).
2	Students will be able to apply control chart techniques to monitor and control process variation.
3	Students will be able to apply patterns on control charts to detect special and common causes of variation.
4	Develop decision-making skills based on control chart outputs for process improvements.

Course Outcome Number	On completing the course, the student will be able to:	PSO Addressed	Blooms Level
CO1	Collect and organize real or simulated process data for control chart analysis.	1, 2,4	Understand and apply
CO2	Construct appropriate control charts ($\bar{X}$-R, p, c, etc.) based on data type and process characteristics.	1, 2,4	Apply
CO3	Detect and interpret assignable causes of variation through chart patterns.	1, 2,4	Analyze and Evaluate
CO4	Use software tools to automate the creation and analysis of control charts.	1,2,4	Understand and Apply

SEMESTER-IV

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		credits
RJSPSTP401	Statistical Computing-IV	Major Practical	Lectures	Practical	
			-	4	2

List of Practical's:

Practical 1: $\overline{X}$ Chart
Practical 2: R Chart
Practical 3: c- Chart
Practical 4: p- Chart
Practical 5: Single Sampling Plan
Practical 6: Double Sampling Plan

M. Sc Statistics Syllabus Semester III & IV

Practical-IV

Research Project

Course code	Course name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPSTARP401	Research Project	Project	Lectures	Practical	
				20	10

Scheme of Examinations

1. Internal Examination 25 marks various modes with different weightage (Presentation, seminar, mcq, quiz etc.)

2. One External (Semester End Examination) of 50 marks. Duration: 2 hours.

3. One Practical at the end of Semester consisting of Practical I 75 marks (based on all three major paper).

4. Minimum marks for passing Semester End Theory and Practical Exam is 40 %.

5. For any KT examinations, there shall be ODD-ODD/EVEN-EVEN pattern followed.

6. A candidate will be allowed to appear for the practical examinations if he/she submits a certified journal of Botany or a certificate from the Head of the department / Institute to the effect that the candidate has completed the practical course of M Sc Semester I Statistics as per the minimum requirements.

7. In case of loss of journal, a candidate must produce a certificate from the Head of the department /Institute that the practical for the academic year were completed by the student. However, such a candidate will be allowed to appear for the practical examination but the marks allotted for the journal will not be granted.

M. Sc Statistics Syllabus Semester III & IV

Paper Pattern for DSC and Elective

Q1.			**Attempt ANY ONE sub question.**	
	A			**15M**
	B			**15M**
Q2.			**Attempt ANY ONE sub question.**	
	A.			**15M**
	B			**15M**
Q3.			**Attempt ANY ONE sub question.**	
	A			**15M**
	B			**15M**
Q4.			**Attempt ANY ONE sub question.**	
	A			**15M**
	B			**15M**